Storage on the Grid – Work Breakdown Structure

Gabriele Garzoglio

v0.1 – Dec 11, 2009
v1.0 – Jan 07, 2011

Introduction

This document presents the work breakdown structure (WBS) of the "Storage on the Grid" project. This project was started as a response to the recorded incidents related to data intensive jobs from Intensity Frontier experiments running on FermiGrid and accessing data from the BlueArc storage (see investigation on stakeholders' usage of storage [5] on Oct 2009). The goal of this project is the evaluation of storage technologies for the use case of data intensive Grid jobs. The storage technologies initially considered are the Hadoop Distributed File System (HDFS) [1], Lustre [2], and Blue Arc (BA) [3]. In Nov 2010, the project has accepted the change request from FermiGrid to evaluate Orange FS [6] as part of this work. The number of technologies for this evaluation was limited to a "small" number, considering the effort available for the project. The technologies were selected at the beginning of the project as an agreement among the collaborators. The targeted infrastructures that will benefit from such evaluation are FermiGrid and the General Physics Computing Farm. This work is lead by the DOCS group in the context of a loose collaboration with groups and departments of the Fermilab Computing Division that expressed interest in the evaluation or have direct expertise in storage: the FermiGrid, OSG Storage, DMS, and FEF groups at Fermilab.

Resources

This plan assumes the availability of the following resources

- Gabriele Garzoglio (marked as GG in the charts): Project Manager. Approx. 40% FTE until Mar 1, 2010 and 20% afterwards. After that date, as per the FY10 / FY11 budgets, Gabriele effort is reduced to 20%.
- Doug Strain and Ted Hesselroth: Developers from the OSG Group. Doug helped at 20% FTE on the project from Mar 15, 2010 to Dec 2010. Doug is marked as NH (New Hire) in the original plans. Ted Hesselroth helps for 20% FTE as of Dec 2010.
- Tanya Levshina (TL): OSG Storage Area Coordinator. We assume the help of her group in the installation of the storage technologies and data movement services (BeStMan, GridFTP, etc.).
- Steve Timm (ST): FermiCloud Project Manager. Steve's group has helped with the deployment of FermiCloud and the integration of the storage solutions with FermiCloud and FermiGrid, our test environment.

- Extra Help (marked as Help! in the charts): Possible help assumed at 50% on limited specific tasks in the phase of planning. Alex Kulyavtsev and Amitoj Singh have provided limited consulting cycles upon request in 2010.

Commented WBS Items and Assumptions

These items were planned in Dec 2009. Their status is revised on Jan 2011.
Additional WBS items are: Orange FS evaluation; Collaboration with HEPiX storage group. See "Plan" section for more information.

1. Document Assessment Process

   1.1. Select relevant storage requirements/metrics – DONE

   See DMS' Lustre evaluation [2]

   1.2. Analyze selected storage metrics for Data Intensive jobs from RunII and IF – DONE[7,8]

   This deliverable provides the baseline of the minimal expected storage performance, given the current *status quo*. Without some external help, this task should be abandoned as it would delay the project of 20 days.

   1.3. Document data access models for the technologies considered – MUTE

   Examples of possible data access models: pre-staged scratch area; tape backed cache; external access mechanism (SRM, GridFTP, …); internal access mechanism (POSIX, SAM, Special API, …); …

2. Deploy the physical and virtual test infrastructure

   2.1. Design HW and VM layout to support the evaluation of storage technologies – DONE

   The deliverable of this task might influence the configuration and design of the FermiCloud infrastructure.

   2.2. Procure / commission Cloud infrastructure – DONE

   This tasks is assumed to be worked on by the FermiCloud team with no effort from this project, except for some initial consultation and design decisions (see previous item). We assume that the infrastructure will be available on Mar 1, 2010.

3. Prepare testing infrastructure – DONE

   This whole activity with its subtasks could benefit from external help. At the lab, several experts have the knowledge of setting up and using storage benchmarks. It is estimated that with their help at 50% FTE, this activity could take 15 calendar days, instead of 27.

   3.1. Commission storage benchmark for technology assessments – DONE

   See DMS' Lustre evaluation. Reuse DMS' storage benchmarks, if possible.

   3.2. Gather and learn to run real user jobs from RunII and IF – DONE

   We assume the availability of physicists from DZero, Minos, CDF, etc. Some people has given already tentatively their availability at the meeting on Nov 19, 2009 [4].

3.3. Develop measurement suite for real user jobs – DONE

4. Assess Lustre

4.1. Deploy storage service and related servers (BeStMan, GridFTP, …) – MUTE

As already tentatively agreed, we assume that the OSG Storage Group will be available for the basic deployment of the storage solutions and interfaces. We assume 3 days of full time effort: this estimate has a potentially large error.

4.2. Integrate FermiGrid with storage service with Lustre – DONE

We assume that the FermiGrid team will be available to work at 0.5 FTE on this task.

4.3. Run benchmarks to compare performance with known results for Lustre. Optimize storage as appropriate – DONE

This task could benefit from external help. The expectation is that most of the work will consist in the study of the metrics and the tuning of the storage parameters. It is estimated that help from a storage expert at 50% FTE could cut down this time from 12 to 5 calendar days.

4.4. Run measurement suite with real jobs for Lustre – 75% DONE

This task could benefit from external help for the same reason as the task above. It is estimated that help from a storage expert at 50% FTE could cut down this time from 12 to 5 calendar days.

4.5. Document Lustre results – 90% DONE

5. Assess HDFS

5.1. Deploy storage service and related servers (BeStMan, GridFTP, …)

As already tentatively agreed, we assume that the OSG Storage Group will be available for the basic deployment of the storage solutions and interfaces. We assume 3 days of full time effort: this estimate has a potentially large error.

5.2. Integrate FG with storage service with HDFS

For example, integrate HDFS VMs as WN. We assume that the FermiGrid team will be available to work at 0.5 FTE on this task.

5.3. Run benchmarks to compare performance with known results for HDFS. Optimize storage as appropriate

This task could benefit from external help for the same reason as task 4.3. It is estimated that help from a storage expert at 50% FTE could cut down this time from 12 to 5 calendar days.

5.4. Run measurement suite with real jobs for HDFS

This task could benefit from external help for the same reason as task 4.3. It is estimated that help from a storage expert at 50% FTE could cut down this time from 12 to 5 calendar days.

5.5. Document HDFS results

6. Assess BA for comparison

To stress the BA from FermiGrid or FermiCloud, we need to generate high IOPS test (e.g. moving a lot of small files), rather than high bandwidth tests (the network might saturate before the BA, in this case). These same tests should also used for Lustre and HDFS, to allow us a direct comparison with BA results.

6.1. Devise minimally disruptive testing technique

For the estimate on the amount of effort (1 FTE week), we assumed the availability of BlueArc experts to help with this task. This plan will define high IOPS tests appropriate for BA.

6.2. Run benchmarks to compare performance with known results for BA. Optimize storage as appropriate

This task could benefit from external help for the same reason as task 4.3. It is estimated that help from a storage expert at 50% FTE could cut down this time from 12 to 5 calendar days.

6.3. Run measurement suite with real jobs for BA

This task could benefit from external help for the same reason as task 4.3. It is estimated that help from a storage expert at 50% FTE could cut down this time from 12 to 5 calendar days.

6.4. Document BA results

7. Reports and documentation

7.1. Relate Data Intensiveness requirements and technology assessment

In both plans we assume that we had external help to deliver the analysis of the data intensiveness requirements for IF and RunII (see task 1.2)

7.2. Document relevance of the study for GPCF

7.3. Study references in literature to assess operational properties, long-term resilience, etc.

7.4. Gather all documentation

Plan

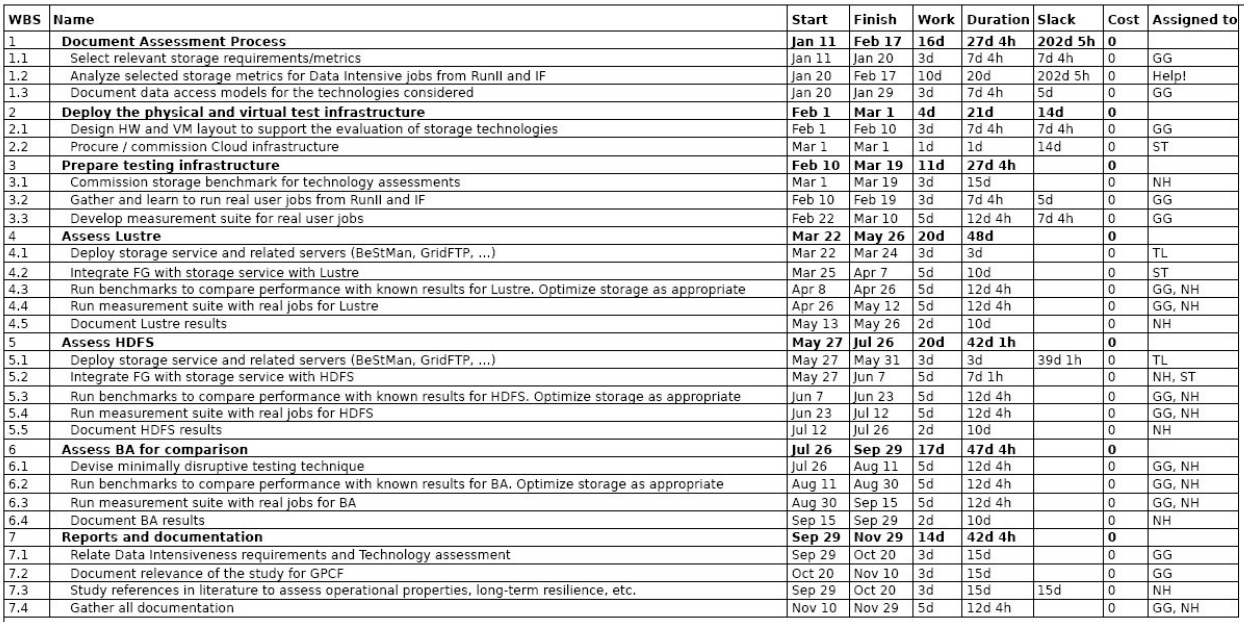
The new estimated completion date of the project is May 2011. The change with respect to the original plan (below) is due to the following reasons:

- FermiCloud was deployed in June 2010, instead of Mar 2010, as assumed.
- The scope was extended to include a collaboration with the HEPiX Storage group. All group reports now include the Nova framework in their studies. This effort took about 4 calendar months, in parallel with other items.
- The time needed for the development of a measurement suite for real user jobs (WBS items 3.2 and 3.3) was underestimated. The project is using the Nova framework in the root-based benchmarking suite. The framework had to be modified to address the needs of the evaluation project. This was unforeseen and took two calendar months, instead of the assumed 20 days.

- The study of the Lustre storage services on FermiCloud and FermiGrid (WBS item 4.2) was extended to include a performance comparison for a deployment of servers on bare metal vs. virtual machines. This almost doubled the amount of work for preparing the test bed and taking the measurements. It also generated a steep learning curve for the team on the deployment and configuration of the virtual machines. This took an additional 4 calendar months of active work in addition to the estimated 2 months.
- Orange FS is an additional technology that the project has accepted to evaluate. We estimate about 1.5 calendar month for the deployment and study.

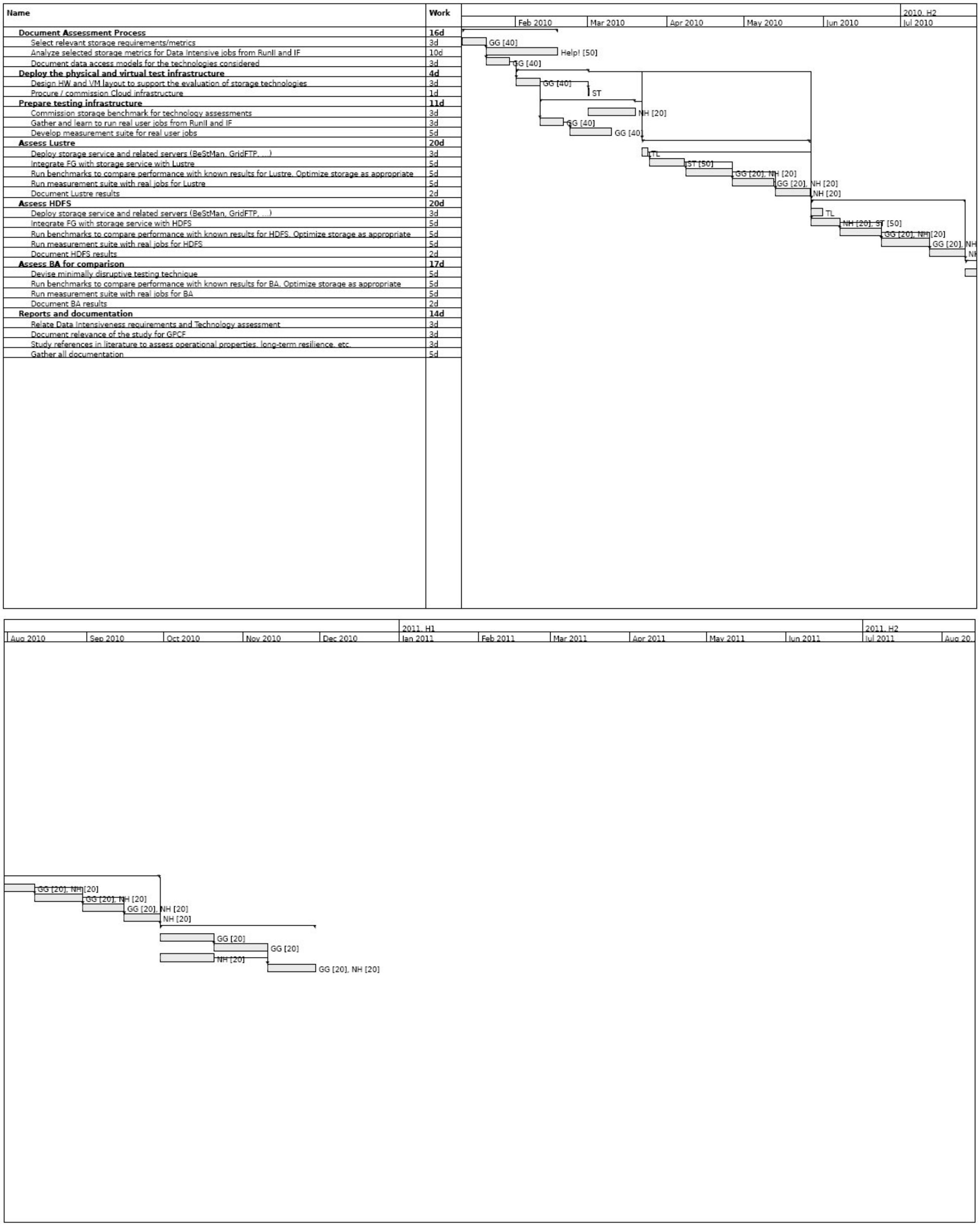
For the record, this is the timeline of the original plan. We still use the plan for an estimate of how long it will take to evaluate Hadoop, BA, and Orange FS.

Original Plans without Help



WBS	Name	Start	Finish	Work	Duration	Slack	Cost	Assigned to
1	**Document Assessment Process**	**Jan 11**	**Feb 17**	**16d**	**27d 4h**	**202d 5h**	**0**	
1.1	Select relevant storage requirements/metrics	Jan 11	Jan 20	3d	7d 4h	7d 4h	0	GG
1.2	Analyze selected storage metrics for Data Intensive jobs from RunII and IF	Jan 20	Feb 17	10d	20d	202d 5h	0	Help!
1.3	Document data access models for the technologies considered	Jan 20	Jan 29	3d	7d 4h	5d	0	GG
2	**Deploy the physical and virtual test infrastructure**	**Feb 1**	**Mar 1**	**4d**	**21d**	**14d**	**0**	
2.1	Design HW and VM layout to support the evaluation of storage technologies	Feb 1	Feb 10	3d	7d 4h	7d 4h	0	GG
2.2	Procure / commission Cloud infrastructure	Mar 1	Mar 1	1d	1d	14d	0	ST
3	**Prepare testing infrastructure**	**Feb 10**	**Mar 19**	**11d**	**27d 4h**		**0**	
3.1	Commission storage benchmark for technology assessments	Mar 1	Mar 19	3d	15d		0	NH
3.2	Gather and learn to run real user jobs from RunII and IF	Feb 10	Feb 19	3d	7d 4h	5d	0	GG
3.3	Develop measurement suite for real user jobs	Feb 22	Mar 10	5d	12d 4h	7d 4h	0	GG
4	**Assess Lustre**	**Mar 22**	**May 26**	**20d**	**48d**		**0**	
4.1	Deploy storage service and related servers (BeStMan, GridFTP, ...)	Mar 22	Mar 24	3d	3d		0	TL
4.2	Integrate FG with storage service with Lustre	Mar 25	Apr 7	5d	10d		0	ST
4.3	Run benchmarks to compare performance with known results for Lustre. Optimize storage as appropriate	Apr 8	Apr 26	5d	12d 4h		0	GG, NH
4.4	Run measurement suite with real jobs for Lustre	Apr 26	May 12	5d	12d 4h		0	GG, NH
4.5	Document Lustre results	May 13	May 26	2d	10d		0	NH
5	**Assess HDFS**	**May 27**	**Jul 26**	**20d**	**42d 1h**		**0**	
5.1	Deploy storage service and related servers (BeStMan, GridFTP, ...)	May 27	May 31	3d	3d	39d 1h	0	TL
5.2	Integrate FG with storage service with HDFS	May 27	Jun 7	5d	7d 1h		0	NH, ST
5.3	Run benchmarks to compare performance with known results for HDFS. Optimize storage as appropriate	Jun 7	Jun 23	5d	12d 4h		0	GG, NH
5.4	Run measurement suite with real jobs for HDFS	Jun 23	Jul 12	5d	12d 4h		0	GG, NH
5.5	Document HDFS results	Jul 12	Jul 26	2d	10d		0	NH
6	**Assess BA for comparison**	**Jul 26**	**Sep 29**	**17d**	**47d 4h**		**0**	
6.1	Devise minimally disruptive testing technique	Jul 26	Aug 11	5d	12d 4h		0	GG, NH
6.2	Run benchmarks to compare performance with known results for BA. Optimize storage as appropriate	Aug 11	Aug 30	5d	12d 4h		0	GG, NH
6.3	Run measurement suite with real jobs for BA	Aug 30	Sep 15	5d	12d 4h		0	GG, NH
6.4	Document BA results	Sep 15	Sep 29	2d	10d		0	NH
7	**Reports and documentation**	**Sep 29**	**Nov 29**	**14d**	**42d 4h**		**0**	
7.1	Relate Data Intensiveness requirements and Technology assessment	Sep 29	Oct 20	3d	15d		0	GG
7.2	Document relevance of the study for GPCF	Oct 20	Nov 10	3d	15d		0	GG
7.3	Study references in literature to assess operational properties, long-term resilience, etc.	Sep 29	Oct 20	3d	15d	15d	0	NH
7.4	Gather all documentation	Nov 10	Nov 29	5d	12d 4h		0	GG, NH

Fig 1: The timeline and resource assignments of the plan ***without help***



Name	Work
Document Assessment Process	**16d**
Select relevant storage requirements/metrics	3d
Analyze selected storage metrics for Data Intensive jobs from RunII and IF	10d
Document data access models for the technologies considered	3d
Deploy the physical and virtual test infrastructure	**4d**
Design HW and VM layout to support the evaluation of storage technologies	3d
Procure / commission Cloud infrastructure	1d
Prepare testing infrastructure	**11d**
Commission storage benchmark for technology assessments	3d
Gather and learn to run real user jobs from RunII and IF	3d
Develop measurement suite for real user jobs	5d
Assess Lustre	**20d**
Deploy storage service and related servers (BeStMan, GridFTP, ...)	3d
Integrate FG with storage service with Lustre	5d
Run benchmarks to compare performance with known results for Lustre. Optimize storage as appropriate	5d
Run measurement suite with real jobs for Lustre	5d
Document Lustre results	2d
Assess HDFS	**20d**
Deploy storage service and related servers (BeStMan, GridFTP, ...)	3d
Integrate FG with storage service with HDFS	5d
Run benchmarks to compare performance with known results for HDFS. Optimize storage as appropriate	5d
Run measurement suite with real jobs for HDFS	5d
Document HDFS results	2d
Assess BA for comparison	**17d**
Devise minimally disruptive testing technique	5d
Run benchmarks to compare performance with known results for BA. Optimize storage as appropriate	5d
Run measurement suite with real jobs for BA	5d
Document BA results	2d
Reports and documentation	**14d**
Relate Data Intensiveness requirements and Technology assessment	3d
Document relevance of the study for GPCF	3d
Study references in literature to assess operational properties, long-term resilience, etc.	3d
Gather all documentation	5d

Fig 2: The Gantt chart for the plan ***without help***

References

[1] G. Attebury, A. Baranovskiy, K. Bloom, B. Bockelman, D. Kcira, J. Letts, T. Levshina, C. Lundestedt, T. Martin, W. Maier, H. Pi, A. Rana, I. Sfiligoi, A. Sim, M. Thomas, F. Wuerthwein, "Hadoop Distributed File System for the Grid", to be published in the proceedings of the IEEE Nuclear Science Symposium, Oct 2009.

[2] G. Oleynik, J. Bakken, R. Rechenmacher, J. Simone, D. Holmgren, N. Seenu, D. Petravick, M. Crawford, S. Fuess, A. Kulyavtsev, D. Litvintsev, A. Moibenko, T. Perelmutov, V. Podstavkov, S. Naymola, S. Wolbers, "Lustre Evaluation as part of 2008 Storage Evaluation", CD-docdb 2817

[3] R. Pasetes, Central NAS Major Upgrade Plan 2009 – doc db 3229

[4] Slides of the Nov 19, 2009 meeting on Storage on FermiGrid and GPCF: docdb 3279

[5] G. Garzoglio, "Storage Services for the Fermilab Grid Facility" – doc db 3279

[6] Orange FS – http://orangefs.net/ – Accessed on Jan 2011

[7] Study of DZero file access metrics – http://home.fnal.gov/~garzogli/storage/dzero-sam-file-access.html – Accessed on Jan 2010

[8] Study of CDF file access metrics – http://home.fnal.gov/~garzogli/storage/cdf-sam-file-access-per-app-family.html – Accessed on Jan 2010